



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2016년05월27일
(11) 등록번호 10-1625124
(24) 등록일자 2016년05월23일

(51) 국제특허분류(Int. Cl.)
G06F 17/30 (2006.01) G06F 17/00 (2006.01)
G06Q 50/00 (2006.01)
(52) CPC특허분류
G06F 17/30011 (2013.01)
G06F 17/00 (2013.01)
(21) 출원번호 10-2015-0013505
(22) 출원일자 2015년01월28일
심사청구일자 2015년01월28일
(56) 선행기술조사문헌
KR1020090069874 A
KR1020140146439 A*
KR1020120123781 A*
전성해, '특허분석을 이용한 지능형시스템의 기술예측', 한국지능시스템학회 논문지 제21권 제1호, 2011, pp.100-105
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
고려대학교 산학협력단
서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)
(72) 발명자
박상성
서울특별시 동대문구 한천로58길 107, 104동 2105호 (이문동, 현대아파트)
전성해
인천광역시 연수구 해돋이로120번길 16, 202동 1702호(송도동, 송도풍림아이원2단지아파트)
(74) 대리인
특허법인충현
(뒷면에 계속)

전체 청구항 수 : 총 5 항

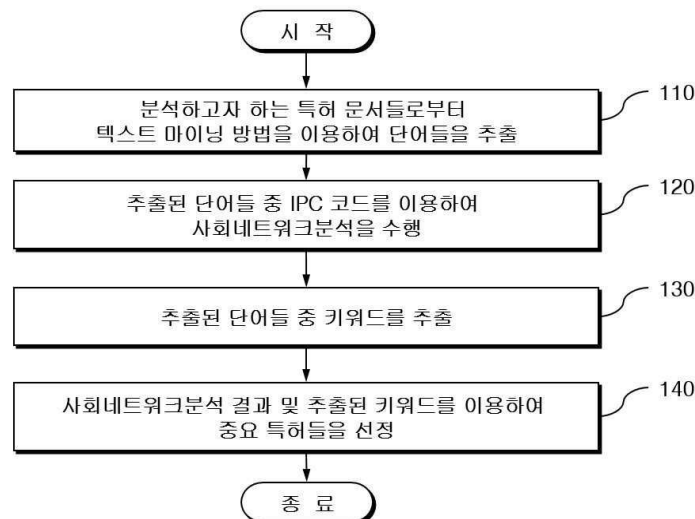
심사관 : 경연정

(54) 발명의 명칭 특허 정량분석을 이용한 기술평가 방법

(57) 요약

본 발명은 특허 문서를 이용하여 기술을 분석하는 방법에 관한 것으로서, 분석하고자 하는 특허 문서들로부터 텍스트 마이닝 방법을 이용하여 단어들을 추출 단계, 상기 추출된 단어들 중 IPC 코드를 이용하여 사회네트워크분석을 수행하는 단계, 상기 추출된 단어들 중 키워드를 추출하는 단계, 및 상기 사회네트워크분석 결과 및 추출된 키워드를 이용하여 중요 특허들을 선정하는 단계를 포함함으로써 객관적이고 정확한 특허 분석이 가능하고, 기술 이전 및 기술 파이낸싱 등에서 리스크를 줄이기 위한 객관적인 기술평가 결과를 제공할 수 있다.

대표도 - 도1



(52) CPC특허분류

G06Q 50/01 (2013.01)

(72) 발명자

장동식

서울특별시 성북구 안암로9길 30, 2동 1102호(안암
동3가, 삼익아파트)

김중찬

경기도 광주시 문화로 112-8 (경안동)

이홍철

서울특별시 서초구 서초대로26길 19, 101동 1202호
(방배동, 브라운스톤방배)

이 발명을 지원한 국가연구개발사업

과제고유번호 2013053572

부처명 교육부

연구관리전문기관 한국연구재단

연구사업명 (이공)일반연구자-기본(유형II)

연구과제명 시맨틱 기술 기반 IP 마이닝 시스템 개발

기 여 율 1/2

주관기관 고려대학교 산학협력단

연구기간 2014.09.01 ~ 2015.08.31

이 발명을 지원한 국가연구개발사업

과제고유번호 F13SN08T4605

부처명 교육부

연구관리전문기관 한국연구재단

연구사업명 BK21PLUS사업

연구과제명 BK21PLUS사업 2차년도 빅데이터운용사업팀 국고지원금

기 여 율 1/2

주관기관 고려대학교 산학협력단

연구기간 2014.03.01 ~ 2015.02.28

명세서

청구범위

청구항 1

특허 문서를 이용하여 기술을 분석하는 방법에 있어서,

특허분석 프로세서가 분석하고자 하는 특허 문서들로부터 텍스트 마이닝 방법을 이용하여 단어들을 추출하는 단계;

상기 특허분석 프로세서가 상기 추출된 단어들 중 IPC 코드를 이용하여 사회네트워크분석을 수행하는 단계;

상기 특허분석 프로세서가 상기 추출된 단어들 중 키워드를 추출하는 단계; 및

상기 특허분석 프로세서가 상기 사회네트워크분석 결과 및 상기 추출된 키워드를 이용하여 중요 특허들을 선정하는 단계를 포함하고,

상기 사회네트워크분석을 수행하는 단계는,

상기 추출된 단어들로부터 IPC 코드들을 추출하는 단계;

상기 추출된 각 IPC 코드들의 근접중심점(Closeness centrality), 중개중심점(Betweenness centrality), 그래프중심점(Graph centrality), 또는 최단거리(shortest distance) 중 하나 이상을 산출하는 단계; 및

상기 산출된 근접중심점, 중개중심점, 그래프중심점, 또는 최단거리 값에 따라 우선순위를 설정하고, 상기 우선순위가 가장 높은 IPC를 중심 IPC 코드로 선택하는 단계를 포함하는 것을 특징으로 하는 방법.

청구항 2

삭제

청구항 3

제 1 항에 있어서,

상기 키워드를 추출하는 단계는,

상기 특허분석 프로세서가 상기 추출된 단어들을 이용하여 문서-단어 행렬을 생성하는 단계;

상기 특허분석 프로세서가 상기 생성된 문서-단어 행렬로부터 키워드를 선택하여 문서-키워드 행렬을 생성하는 단계;

상기 특허분석 프로세서가 상기 문서-키워드 행렬을 이용하여 회귀모델을 생성하는 단계;

상기 특허분석 프로세서가 상기 생성된 회귀모델의 매개변수들 중 제 1 임계치 이하의 유의확률(p-value)을 갖는 매개변수에 해당하는 키워드를 선택하는 단계; 및

상기 특허분석 프로세서가 상기 선택된 키워드 간의 유의확률을 이용하여 중요 키워드를 도출하는 단계를 포함하는 방법.

청구항 4

제 1 항에 있어서,

상기 키워드를 추출하는 단계는,

상기 특허분석 프로세서가 상기 추출된 단어들을 이용하여 문서-단어 행렬을 생성하는 단계;

상기 특허분석 프로세서가 상기 생성된 문서-단어 행렬로부터 키워드를 선택하여 문서-키워드 행렬을 생성하는 단계;

상기 특허분석 프로세서가 상기 문서-키워드 행렬을 이용하여 키워드를 특징으로 하는 k-means 분석을 통해 특

히 문서들을 기술군집으로 군집화하는 단계;

상기 특허분석 프로세서가 상기 군집화된 기술군집의 키워드들을 이용하여 각 기술군집의 대표기술을 도출하는 단계; 및

상기 특허분석 프로세서가 중요 키워드를 가장 많이 포함하는 기술군집을 중요 기술군집으로 도출하는 단계를 포함하는 방법.

청구항 5

제 1 항에 있어서,

상기 중요 특허들을 선정하는 단계는,

상기 특허분석 프로세서가 출원 후 경과기간, 청구항 수, IPC 코드의 수, 중요 키워드의 포함 수, 또는 사회네트워크분석 결과 중 하나 이상의 결과 값과 상기 결과 값의 평균을 산출하는 단계; 및

상기 특허분석 프로세서가 상기 산출된 결과 값의 평균으로부터 기술가치 순위를 선정하는 단계를 포함하는 방법.

청구항 6

제 1 항에 있어서,

상기 특허분석 프로세서가 상기 사회네트워크분석 결과, 상기 추출된 키워드, 또는 상기 선정된 중요 특허들 중 하나 이상을 사용자에게 제공하는 단계를 더 포함하는 방법.

발명의 설명

기술 분야

[0001] 본 발명은 특허를 분석하는 방법에 관한 것으로서, 보다 구체적으로, 사회네트워크분석 및 키워드 분석을 통해 기술평가를 수행하는 방법에 관한 것이다.

배경 기술

[0002] 기술은 사람들의 삶의 질을 향상시키기 위한 활동이나 지식을 뜻하므로 기술의 효율적인 관리는 매우 중요하다고 할 수 있다. 기술경영은 비즈니스와 산업 부문에서 중요한 역할을 하고 있다. 기술경영은 기술평가, 신제품 개발, 기술 혁신, R&D 경영, 기술예측과 같은 많은 연구들을 포함하고 있다. 그 중 기술평가는 기술을 지적 재산으로써 가치를 평가하는 것을 의미한다. 기술의 가치 평가 기준은 개발자나 평가자의 관점에 따라 달라질 수 있다. 최근 기술평가의 결과가 기술 마케팅 또는 기술이전 및 라이선스, R&D계획, 기술 투자에 이용되기 때문에 체계적이고 실용적인 기술평가의 중요성이 강조되고 있다. 종래의 기술은 해당 분야 전문가에 지식과 경험에 의해 만들어진 다양한 지표들을 이용해 기술을 평가하였고 이 지표들은 기술의 경제적 가치에 중점을 두고 있었다. 그러나 기술은 불확실성을 포함한 무형의 특성을 갖고 있기 때문에 미래의 경제적 가치를 완벽히 예측하는 것이 불가능하다. 그러므로 경제적 지표를 이용한 기술 평가는 어려움이 있다. 경제적 지표를 이용한 평가 방법 외로 전문가와의 의견과 견해를 통한 기술평가 방법도 있었다. 이러한 종래의 기술들은 전문가의 지식과 경험에 의존한 기술평가의 결과물은 일관성을 보장하지 못하고 주관적이다. 그러나 기술 이전 또는 마케팅을 위해 기술평가는 반드시 필요한 기술이다.

선행기술문헌

특허문헌

[0003] (특허문헌 0001) 한국공개특허 "조사 대상 문서의 문서 특징 분석 장치(10-2006-0095565)"

발명의 내용

해결하려는 과제

[0004] 본 발명이 해결하고자 하는 과제는 사회네트워크분석 및 키워드 분석을 통해 기술평가를 수행하는 방법을 제공하는 것이다.

과제의 해결 수단

[0005] 본 발명은 상기 과제를 해결하기 위하여, 특허 문서를 이용하여 기술을 분석하는 방법에 있어서, 분석하고자 하는 특허 문서들로부터 텍스트 마이닝 방법을 이용하여 단어들을 추출하는 단계; 상기 추출된 단어들 중 IPC 코드를 이용하여 사회네트워크분석을 수행하는 단계; 상기 추출된 단어들 중 키워드를 추출하는 단계; 및 상기 사회네트워크분석 결과 및 상기 추출된 키워드를 이용하여 중요 특허들을 선정하는 단계를 포함하는 방법을 제공한다.

[0006] 본 발명의 다른 실시예에 의하면, 상기 사회네트워크분석을 수행하는 단계는, 상기 추출된 단어들로부터 IPC 코드들을 추출하는 단계; 상기 추출된 각 IPC 코드들의 근접중심점(Closeness centrality), 중개중심점(Betweenness centrality), 그래프중심점(Graph centrality), 또는 최단거리(shortest distance) 중 하나 이상을 산출하는 단계; 및 상기 산출된 근접중심점(Closeness centrality), 중개중심점(Betweenness centrality), 그래프중심점(Graph centrality), 또는 최단거리(shortest distance) 값에 따라 우선순위를 설정하고, 상기 우선순위가 가장 높은 IPC를 중심 IPC 코드로 선택하는 단계를 포함하는 방법일 수 있다.

[0007] 본 발명의 다른 실시예에 의하면, 상기 키워드를 추출하는 단계는, 상기 추출된 단어들을 이용하여 문서-단어 행렬을 생성하는 단계; 상기 생성된 문서-단어 행렬로부터 키워드를 선택하여 문서-키워드 행렬을 생성하는 단계; 상기 문서-키워드 행렬을 이용하여 회귀모델을 생성하는 단계; 상기 생성된 회귀모델의 매개변수들 중 제 1 임계치 이하의 유의확률(p-value)을 갖는 매개변수에 해당하는 키워드를 선택하는 단계; 및 상기 선택된 키워드 간의 유의확률을 이용하여 중요 키워드를 도출하는 단계를 포함하는 방법일 수 있다.

[0008] 본 발명의 다른 실시예에 의하면, 상기 키워드를 추출하는 단계는, 상기 추출된 단어들을 이용하여 문서-단어 행렬을 생성하는 단계; 상기 생성된 문서-단어 행렬로부터 키워드를 선택하여 문서-키워드 행렬을 생성하는 단계; 상기 문서-키워드 행렬을 이용하여 키워드를 특징으로 하는 k-means 분석을 통해 특허 문서들을 기술군집으로 군집화하는 단계; 상기 군집화된 기술군집의 키워드들을 이용하여 각 기술군집의 대표기술을 도출하는 단계; 및 중요 키워드를 가장 많이 포함하는 기술군집을 중요 기술군집으로 도출하는 단계를 포함하는 방법일 수 있다.

[0009] 본 발명의 다른 실시예에 의하면, 상기 중요 특허들을 선정하는 단계는, 출원 후 경과기간, 청구항 수, IPC 코드의 수, 중요 키워드의 포함 수, 또는 사회네트워크분석 결과 중 하나 이상의 결과 값과 상기 결과 값의 평균을 산출하는 단계; 및 상기 산출된 결과 값의 평균으로부터 기술가치 순위를 선정하는 단계를 포함하는 방법일 수 있다.

[0010] 본 발명의 다른 실시예에 의하면, 상기 사회네트워크분석 결과, 상기 추출된 키워드, 또는 상기 선정된 중요 특허들 중 하나 이상을 사용자에게 제공하는 단계를 더 포함하는 방법일 수 있다.

발명의 효과

[0011] 본 발명에 따르면, 신속하고 정확하며 보다 객관적으로 결과를 제시하여 기업의 R&D계획 및 경영전략을 수립하는데 있어 보다 객관적인 정보를 제공한다. 또한, 기술 이전 및 기술 파이낸싱 등에서 리스크를 줄이기 위한 객관적인 기술평가 결과를 제공할 수 있다.

도면의 간단한 설명

[0012] 도 1은 본 발명의 일 실시예에 따른 특허분석방법의 흐름도이다.

도 2 내지 도 6은 본 발명의 실시예에 따른 특허분석방법의 흐름도이다.

도 7 내지 도 10은 본 발명의 실시예에 따른 특허분석의 과정을 나타낸 것이다.

발명을 실시하기 위한 구체적인 내용

[0013] 본 발명에 관한 구체적인 내용의 설명에 앞서 이해의 편의를 위해 본 발명이 해결하고자 하는 과제의 해결 방안

의 개요 혹은 기술적 사상의 핵심을 우선 제시한다.

- [0014] 본 발명의 일 실시예에 따른 특허 문서를 이용하여 기술을 분석하는 방법에 있어서, 분석하고자 하는 특허 문서들로부터 텍스트 마이닝 방법을 이용하여 단어들을 추출하는 단계, 상기 추출된 단어들 중 IPC 코드를 이용하여 사회네트워크분석을 수행하는 단계, 상기 추출된 단어들 중 키워드를 추출하는 단계, 및 상기 사회네트워크분석 결과 및 상기 추출된 키워드를 이용하여 중요 특허들을 선정하는 단계를 포함한다.
- [0015] 이하 첨부된 도면을 참조하여 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자가 본 발명을 용이하게 실시할 수 있는 실시 예를 상세히 설명한다. 그러나 이들 실시예는 본 발명을 보다 구체적으로 설명하기 위한 것으로, 본 발명의 범위가 이에 의하여 제한되지 않는다는 것은 당업계의 통상의 지식을 가진 자에게 자명할 것이다.
- [0016] 본 발명이 해결하고자 하는 과제의 해결 방안을 명확하게 하기 위한 발명의 구성을 본 발명의 바람직한 실시예에 근거하여 첨부 도면을 참조하여 상세히 설명하되, 당해 도면에 대한 설명시 필요한 경우 다른 도면의 구성요소를 인용할 수 있음을 미리 밝혀둔다. 아울러 본 발명의 바람직한 실시 예에 대한 동작 원리를 상세하게 설명함에 있어 본 발명과 관련된 공지 기능 혹은 구성에 대한 구체적인 설명 그리고 그 이외의 제반 사항이 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우, 그 상세한 설명을 생략한다.
- [0017] 도 1은 본 발명의 일 실시예에 따른 특허분석방법의 흐름도이다.
- [0018] 본 발명의 일 실시예에 따른 특허 문서를 이용하여 기술을 분석하는 방법은 다음의 일련의 과정을 통해 구현된다.
- [0019] 110 단계는 분석하고자 하는 특허 문서들로부터 텍스트 마이닝 방법을 이용하여 단어들을 추출하는 단계이다.
- [0020] 보다 구체적으로, 분석하고자 하는 문서들로부터 텍스트 마이닝 방법을 이용하여 단어들을 추출한다. 텍스트 마이닝(Text Mining) 방법은 자연어로 구성된 비정형 텍스트 데이터에서 패턴 또는 관계를 추출하여 가치와 의미 있는 정보를 찾아내는 마이닝 기법으로, 사람들이 말하는 언어를 이해할 수 있는 자연어처리 기술에 기반한 기술이다. 상기 단어들은 문장분석(parsing) 및 말뭉치분석(corpus)를 이용하여 추출할 수 있다.
- [0021] 상기 분석하고자 하는 문서들은 텍스트 형식의 문서들로, 특히 특허문서일 수 있다. 특허문서에 대한 단어를 추출함에 있어서, 유의한 단어들을 추출하는 정확성 및 효율성을 위하여, 특허문서의 발명의 명칭, 요약, 특허청구범위, 또는 발명의 상세한 설명 중 하나 이상의 부분에서 단어들을 추출할 수 있다. 특허문서의 발명의 명칭, 요약, 특허청구범위에는 해당 특허문서에서 다루고자 하는 발명의 핵심적인 부분이 포함되고 해당 발명의 상세한 내용은 발명의 상세한 설명에 포함되는바, 특허문서의 발명의 명칭, 요약, 특허청구범위, 또는 발명의 상세한 설명 중 하나 이상의 부분에서 단어들을 추출할 수 있다.
- [0022] 120 단계는 상기 추출된 단어들 중 IPC 코드를 이용하여 사회네트워크분석을 수행하는 단계이다.
- [0023] 보다 구체적으로, 110 단계에서 추출된 단어들 중 IPC 코드를 이용하여 사회네트워크 분석을 수행한다. IPC 코드는 국제특허 분류 기호로써 특허문서를 체계적으로 분류하고자 WIPO에서 만든 분류기호이다. IPC 코드를 참고한다면 특허문서를 모두 읽지 않더라도 기술 분류를 가능할 수 있다. 따라서, IPC 코드를 이용하여 기술을 분석하고, 대표기술을 선정한다면, 신속하고 정확하게 해당 분야의 중요 IPC 코드를 확인할 수 있다. IPC 코드뿐만 아니라 키워드를 이용하여 사회네트워크분석을 수행하여 중요 키워드를 확인할 수도 있다.
- [0024] 중요 IPC 코드를 선정하기 위하여, IPC 코드들의 관계에 대한 사회 네트워크 분석을 수행한다. 사회 네트워크 분석(Social Network Analysis, SNA)은 점과 선으로 구성된 네트워크를 사회과학적 또는 통계적인 관점에서 분석한 것이다. 이전에는 사회과학 문제를 해결하기 위하여 사용되었으며 현재에는 마케팅, 고객관계관리 등 많은 영역에서 활용되고 있다. 기술과 기술 간의 관계를 이해하고 분석하여 대표기술을 신속하고 정확하게 찾을 수 있도록 사회 네트워크 분석을 이용한다. 사회 네트워크 분석에서는 vertex (or node)와 edge (or connection) 또는 edge들의 분석에 의한 vertex들 사이의 관계설명 등 총 2가지로 구성되어있다. IPC 코드 간의 관계를 시각화할 수 있어, 이를 통해 쉽게 분석이 가능해진다.
- [0025] 120 단계에 대한 구체적인 과정은 도 2의 210 단계 내지 230 단계에 의해 구현될 수 있다.
- [0026] 210 단계는 상기 추출된 단어들로부터 IPC 코드들을 추출하는 단계이다.

[0027] 보다 구체적으로, 110 단계에서 추출된 단어들 중 IPC 코드들을 추출한다. IPC 코드는 4개의 등급(1: Section, 2: Class, 3: Subclass, 4: group)으로 구성된다. IPC 코드의 4번째 등급까지 이용할 경우 너무 세세하게 분류될 수 있는바, Subclass까지만 이용할 수 있다.

[0028] 220 단계는 상기 추출된 각 IPC 코드들의 근접중심점(Closeness centrality), 중개중심점(Betweenness centrality), 그래프중심점(Graph centrality), 또는 최단거리(shortest distance) 중 하나 이상을 산출하는 단계이다.

[0029] 보다 구체적으로, 사회 네트워크 분석을 수행한 후, 각 IPC 코드의 접중심점(CC), 중개중심점(BC), 그래프 중심점(GC), 최단거리를 중요 IPC 코드의 선정을 위한 기준으로 이용할 수 있다.

[0030] 상기 중요 IPC 코드를 선정하는데 이용되는 기준으로 근접중심점(Closeness centrality), 중개중심점(Betweenness centrality), 그래프중심점(Graph centrality), 또는 최단거리(shortest distance) 중 하나 이상을 이용하며, 아래와 같이 산출될 수 있다.

[0031] 근접중심점은 아래와 같이 정의할 수 있다.

수학식 1

$$CC(i) = \frac{n-1}{\sum_{i \neq j} sdist(i,j)}, i, j = 1, 2, \dots, n$$

[0032]

[0033] 수학식 1에서 sdist(i,j)는 vertex i와 j사이의 가장 짧은 거리이다. 따라서 중요 IPC 코드 선정을 위해서는 가장 큰 CC값을 갖고 있는 vertex를 선정한다. 다음으로 중개중심성은 다음과 같이 나타낼 수 있다.

수학식 2

$$BC(i) = \sum_{i \neq k, i \neq j} \sum_{k \neq j} \frac{sdist(i,j,k)}{sdist(i,j)}, i, j, k = 1, 2, \dots, n$$

[0034]

[0035] 수학식 2에서 sdist(i,j,k)는 k를 통한 i부터 j사이의 가장 짧은 거리를 나타낸다. 중요 IPC 코드를 선정하기 위해서는 CC와 같이 BC가 가장 큰 값을 갖는 vertex를 선택한다. 그래프 중심점은 다음 수학식 3과 같이 나타내며 CC, BC와 같이 GC값이 가장 큰 vertex를 중요 IPC 코드로 선정한다.

수학식 3

$$GC(i) = \frac{1}{Max_{i \neq j} sdist(i,j)}, i, j = 1, 2, \dots, n$$

[0036]

[0037] Degree measure은 각 vertex에 연결된 connection의 개수를 나타낸다.

[0038] 230 단계는 산출된 근접중심점(Closeness centrality), 중개중심점(Betweenness centrality), 그래프중심점(Graph centrality), 또는 최단거리(shortest distance) 값에 따라 우선순위를 설정하고, 상기 우선순위가 가장 높은 IPC를 중심 IPC 코드로 선택하는 단계이다.

[0039] 보다 구체적으로, 220 단계에서 산출된 근접중심점(Closeness centrality), 중개중심점(Betweenness centrality), 그래프중심점(Graph centrality), 또는 최단거리(shortest distance) 값에 따라 우선순위를 설정하고, 상기 우선순위가 가장 높은 IPC를 중심 IPC 코드로 선택한다.

[0040] 도 2(a)과 같은 SNA 그래프가 있다고 가정할 때, degree measure은 다음 표 1과 같이 나타낼 수 있다. 표 1에서

T1의 degree가 가장 많으므로 T1을 중요 IPC 코드로 선정할 수 있다.

표 1

vertwx	degree
T1	3
T2	2
T3	1
T4	2

예를 들어, 상기 4가지 기준 모두를 이용하여 중요 IPC 코드를 선정하는 경우, 아래와 같이, 각 값들을 계산하여 표로 표현할 수 있다.

표 2

IPC 코드	근접중심점	중개중심점	그래프 중심점	최근접거리
A01A	0.56	1.61	0.5	8
A02B	0.75	8.25	0.33	7
B01C	0.90	9.73	0.5	2
B02D	0.81	11.0	0.5	3

표 2에서 4개의 IPC코드 중 B01C가 0.90으로 근접중심점(Closeness Centrality) 값이 가장 높다. 다음으로 중개중심점(Betweenness Centrality)에서는 11.0으로 B02D가 가장 높다. 그래프 중심점(Graph Centrality)의 경우 0.5로 A01A, B01C, 및 B02D가 같으며 마지막으로 최근접거리(Shortest Distance)의 경우 B01C가 2로 가장 짧고 다음으로는 B02D가 3으로 짧다. 따라서 B01C이 다른 IPC코드들과 연결되어 있는 길이가 그 외의 서로 다른 IPC 코드들 사이의 연결된 길이가 가장 짧으므로 최종적으로 B01C가 중요 IPC 코드임을 알 수 있다.

중요 IPC 코드를 선택하기 위하여, 상기 산출된 근접중심점(Closeness centrality), 중개중심점(Betweenness centrality), 그래프중심점(Graph centrality), 또는 최근접거리(shortest distance)에 가중치를 부여하여 상기 우선순위를 설정할 수 있다. 두 개 이상의 기준 값을 이용하여 IPC 코드간 우선순위를 설정함에 있어서, 기준 값들 간의 중요도에 따라 가중치를 다르게 부여할 수 있다. 예를 들어, 최근접거리에 최고 가중치를 두어, 다른 값들에 비해 중심 IPC 코드 선택에 중요성을 가지도록 할 수 있다.

130 단계는 상기 추출된 단어들 중 키워드를 추출하는 단계이다.

보다 구체적으로, 110 단계에서 추출된 단어들 중 키워드를 추출한다. 키워드는 해당 문서가 포함하고 있는 핵심 단어로서 어떤 키워드를 포함하고 있는지를 확인하여 중요 특허인지를 확인할 수 있다. 중요 키워드를 분석함에 있어서, 120 단계의 사회네트워크분석 방법을 이용할 수도 있다.

130 단계에 대한 구체적인 과정은 도 3의 310 단계 내지 350 단계 또는 도 4의 410 단계 내지 450 단계에 의해 구현될 수 있다.

310 단계는 상기 추출된 단어들을 이용하여 문서-단어 행렬을 생성하는 단계이다.

보다 구체적으로, 분석한 문서와 추출된 단어들을 이용하여 문서-단어 행렬(Document-Term Matrix, DTM)을 생성한다. Corpus와 텍스트 마이닝 기법의 text repository를 사용하여 문서-단어 행렬을 구성한다. 문서-단어 행렬은 문서와 단어 간의 관계를 나타낸 행렬로, 문서에 각 단어가 나타나는 빈도수를 나타낸다. 행렬의 행과 열은 단어와 문서들로 구성된다. 각 원소는 각 문서에서 단어 발생 빈도 값이다.

320 단계는, 상기 생성된 문서-단어 행렬로부터 키워드를 선택하여 문서-키워드 행렬을 생성하는 단계이다.

보다 구체적으로, 상기 단어들 중 키워드를 선택하여 문서-키워드 행렬을 생성한다. 단어들 중 키워드를 선택함으로써 행의 수를 줄여 빠르고 정확한 분석을 가능하도록 한다.

문서-키워드 행렬을 생성함에 있어서, 문서-단어 행렬로부터 제 4 임계치 이상의 발생 빈도 값을 갖는 단어들을 선정할 수 있다. 상기 제 4 임계치는 미리 설정되어 있거나, 문서-단어 행렬의 행의 수에 따라 결정될 수 있다. 발생 빈도가 높을수록 키워드일 확률이 높을바, 발생 빈도 값에 따라 단어를 선정할 수 있다. 선정된 단어들 중 기술과 관련이 없는 단어들을 제거하여 문서-키워드 행렬을 생성할 수 있다. 발생 빈도는 높으나, 기술과 관련

이 없는 단어들을 제거함으로써 정확성을 높일 수 있다. 예를 들어, "is", "the"와 같이 의미 없는 단어들을 제거할 수 있다. 제거되는 단어들을 미리 설정되어 있거나 사용자의 설정 입력에 따라 특정 단어들이 제거될 수 있다.

[0054] 330 단계는 상기 문서-키워드 행렬을 이용하여 회귀모델을 생성하는 단계이다.

[0055] 보다 구체적으로, 문서-키워드 행렬을 이용하여 회귀모델을 생성한다. 회귀모델은 하나 또는 그 이상의 독립변수의 종속변수에 대한 영향의 추정을 할 수 있는 통계기법으로 키워드의 통계적 분석이 가능하다. 즉, 회귀모델을 생성하여 선택된 키워드들 중 통계적으로 유의한 키워드를 찾을 수 있다. 키워드 선택은 통계적으로 유의한 분석이 아닌바, 회귀모델을 이용함으로써 통계적으로 유의한 분석이 가능하다. 상기 문서-키워드 행렬로부터 회귀모델을 생성한다. 회귀모델은 다음과 같이 나타낼 수 있다.

수학식 4

$$Z = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + \epsilon$$

[0056]

[0057] Z는 종속변수이고, X_s 는 독립변수이다. 종속변수는 영향을 받는 기술이고, 독립변수는 개발된 기술들을 의미한다. 즉, 종속변수는 분석을 하고자 하는 기술이고, 독립변수는 해당 기술에 관련하여 개발된 기술들로 문서분석을 통해 도출되는 키워드들에 해당한다. ϵ 는 오차항이다. 회귀 매개변수 β_k 는 두 기술들 사이의 인과관계의 강도를 나타낸다. Full regression model에서는 아래와 같이, 모든 독립변수들을 사용하고 가설을 세운다.

수학식 5

$$H_0 : \beta_i = 0 \text{ vs. } H_1 : \beta_i \neq 0, \quad i = 1, 2, \dots, k$$

[0058]

[0059] Z 기술에 영향을 미치는 X 기술들을 찾기 위해, 가설검정을 실시할 수 있다. 귀무가설 H_0 는 i 번째 회귀 매개변수가 0이라는 것을 나타낸다. 이 의미는 i 번째 기술은 기술 Z에 영향을 미치지 않는다는 것이다. 또한 대립가설 H_1 은 i 번째 회귀변수가 0이 아니라는 것을 나타낸다. 즉, 기술 Z는 기술 X_i 에 의존하고 있다는 것이다. 기술 X_i 가 Z에 유의적인 영향을 미치는지 판단하기 위해, 통계적 검정을 통하여 H_0 를 거절할 수 있다. 가설검정을 위해 자유도가 $n-(k+1)$ 인 t분포를 사용한다. n과 k는 데이터 크기와 변수의 개수이다. 만약 H_0 가 참이라면, 아래와 같이 검정통계량을 계산할 수 있다.

수학식 6

$$t = \frac{\beta}{S_{\beta}}$$

[0060]

[0061] $\hat{\beta}_i$ 는 β_i 의 추정치이고, $S_{\hat{\beta}_i}$ 는 $\hat{\beta}_i$ 의 표준오차이다. H_0 를 거절하기 위해서는 다음조건을 만족시켜야 한다.

수학식 7

$$|t| > t_{(n-(k+1)), \alpha/2}$$

[0062]

[0063] α 는 유의수준이고, 또한 t 분포표를 이용하여 $t_{(n-(k+1)), \alpha/2}$ 의 값을 얻을 수 있다.

[0064] 340 단계는 상기 생성된 회귀모델의 매개변수들 중 제 1 임계치 이하의 유의확률(p-value)을 갖는 매개변수에

해당하는 키워드를 선택하는 단계이다.

[0065] 보다 구체적으로, 빠르고 정확한 분석을 위하여, 유의한 키워드를 선택하여 키워드의 수를 줄일 수 있다. 이를 위하여, 330 단계에서 생성된 회귀모델의 매개변수들 중 제 1 임계치 이하의 유의확률(p-value)을 갖는 매개변수에 해당하는 키워드를 선택할 수 있다. 유의확률은 매개변수가 유의한지 안 한지 확인하기 위해 요구되는 가장 작은 유의수준으로, 만약 매개변수의 회귀결과에서 유의확률이 0.05보다 작다면 이 매개변수는 유의한 것으로 판단할 수 있다. 점정통계량의 유의확률이 0.05보다 작을 때, X_i 의 기술은 유의하다고 판단한다. 즉, 축소된 회귀모델을 생성하기 위하여, 유의확률이 0.05보다 작은 매개변수에 해당하는 키워드들을 선택한다. 핵심 키워드를 선정함으로써 정확도와 속도 면에서 매우 효율적이다.

[0066] 350 단계는 상기 선택된 키워드 간의 유의확률을 이용하여 중요 키워드를 도출하는 단계이다.

[0067] 보다 구체적으로, 상기 선택된 유의한 키워드들을 분석하고자 하는 기술과의 관계를 키워드 간의 유의확률을 이용하여 도출할 수 있다. 상기 키워드 간의 관계는 상기 키워드 간의 유의확률을 이용하여 계층을 나누어 나타낼 수 있다. 계층을 나누어, 기술 계층 다이어그램(A Hierarchical Diagram of Technology, HDT)으로 나타낼 수 있다. 상위 계층 기술과 하위 계층의 기술에 대한 관계는 다음과 같이 나타낼 수 있다.

수학적식 8

$$[0068] \quad X_1 = u_{01} + u_{11} W_1$$

[0069] X_1 은 영향을 받은 기술이고, W_1 은 개발된 기술이다. u_{11} 은 X_1 과 W_1 의 인과관계 강도이다. 상기 모델은 기술 Z를 예측하기 위해 모든 인과관계 강도를 찾는 것이다. W_m 기술은 $(ump*vpn*rn)$ 에 인과관계 강도에 따라 기술Z 개발에 영향을 미친다. 그러므로 TF의 Z를 제외하고 모든 기술의 모든 인과관계 강도를 산출할 수 있다.

[0070] 중요 키워드를 도출하기 위해, 상기 선택된 키워드 중 제 2 임계치 이하의 유의확률을 갖는 매개변수에 해당하는 하나 이상의 키워드를 중요 키워드로 선정할 수 있다. 제 1 임계치 이하의 유의확률을 갖는 매개변수에 해당하는 키워드들 중 제 2 임계치 이하의 유의확률을 갖는 매개변수에 해당하는 하나 이상의 키워드를 중요 키워드로 선정한다. 중요 키워드는 분석하고자 하는 기술과 가장 관련이 높은 기술들로, 제 2 임계치 이하의 유의확률을 갖는 매개변수에 해당하는 키워드를 상위계층에 해당하는 중요 키워드로 선정한다. 제 2 임계치는 미리 설정되어 있거나, 사용자의 입력에 따라 설정될 수 있다. 또는 유의확률의 낮은 순서대로 미리 설정된 수만큼 중요 키워드로 선정할 수도 있다.

[0071] 상기 중요 키워드와 중요 키워드 이외의 키워드들 간의 회귀모델 분석을 통해, 제 3 임계치 이하의 유의확률을 갖는 매개변수에 해당하는 키워드를 상기 중요 키워드와 관계성이 있는 키워드로 선정할 수도 있다.

[0072] 보다 구체적으로, 중요 키워드를 선정한 후 다음 계층의 키워드를 선정하는 단계이다. 계층의 깊이는 미리 설정되어 있거나, 키워드의 수, 또는 분석하고자 하는 정도에 따라 달라질 수 있다. 중요 키워드와 관련된 다음 계층의 키워드를 선정하기 위하여, 중요 키워드와 중요 키워드 이외의 키워드들 간의 회귀모델 분석을 수행한다. 상기 중요 키워드와 중요 키워드 이외의 키워드들 간의 회귀모델 분석을 통해, 제 3 임계치 이하의 유의확률을 갖는 매개변수에 해당하는 키워드를 상기 중요 키워드와 관계성이 있는 키워드로 선정한다.

[0073] 이하 도 4의 410 단계 내지 450 단계에 대해 설명하도록 한다.

[0074] 410 단계는 추출된 단어들을 이용하여 문서-단어 행렬을 생성하는 단계이고, 420 단계는 상기 생성된 문서-단어 행렬로부터 키워드를 선택하여 문서-키워드 행렬을 생성하는 단계이다. 410 단계 내지 420 단계는 도 3의 310 단계 내지 320 단계에 대응한다.

[0075] 430 단계는 상기 문서-키워드 행렬을 이용하여 키워드를 특징으로 하는 k-means 분석을 통해 특허 문서들을 기술군집으로 군집화하는 단계이다.

[0076] 보다 구체적으로, 키워드들을 이용하여 문서들이 어떤 형태로 분류될 수 있는지를 확인하기 위하여, k-means 분석을 이용할 수 있다. 이를 통해 특허 문서들을 기술군집으로 군집화할 수 있다. k-means 분석은 주어진 데이터를 k 개의 클러스터로 묶는 방법으로 각 클러스터와 거리차이의 분산을 최소화하는 방식으로 수행된다. 실루엣 측도를 이용하여 최대 평균 실루엣 넓이를 구하는 k를 설정할 수도 있다. 이는 미리 설정되거나, 특허문서의 양

에 따라 달라질 수 있다.

- [0077] 440 단계는 상기 군집화된 기술군집의 키워드들을 이용하여 각 기술군집의 대표기술을 도출하는 단계이다.
- [0078] 보다 구체적으로, 군집화된 기술군집의 키워드들을 분석하여 각 기술군집의 대표기술을 도출할 수 있다.
- [0079] 450 단계는 중요 키워드를 가장 많이 포함하는 기술군집을 중요 기술군집으로 도출하는 단계를 포함하는 방법.
- [0080] 보다 구체적으로, 문서-키워드 행렬을 통해 도출되는 중요 키워드를 가장 많이 포함하는 기술군집을 중요 기술군집으로 도출할 수 있다. 포함된 중요 키워드의 수에 따라 기술군집의 우선순위를 산출하고, 그에 따라 중요 기술군집을 도출한다.
- [0081] 140 단계는 상기 사회네트워크분석 결과 및 상기 추출된 키워드를 이용하여 중요 특허들을 선정하는 단계이다.
- [0082] 보다 구체적으로, 120 단계의 결과 및 130 단계의 결과를 통해 중요 특허를 객관적으로 선정할 수 있다.
- [0083] 140 단계에 대한 구체적인 과정은 도 5의 510 단계 내지 520 단계에 의해 구현될 수 있다.
- [0084] 510 단계는 출원 후 경과기간, 청구항 수, IPC 코드의 수, 중요 키워드의 포함 수, 또는 사회네트워크분석 결과 중 하나 이상의 결과 값과 상기 결과 값의 평균을 산출하는 단계이다.
- [0085] 보다 구체적으로, 110 단계에서 추출된 단어들로부터 추출할 수 있는 출원 후 경과기간, 청구항 수, IPC 코드의 수, 120 단계의 결과로부터 도출되는 군집중심점, 중개중심점, 그래프중심점, 또는 최근접거리, 130 단계의 결과로부터 도출되는 중요 키워드의 포함 수 중 하나 이상의 결과 값을 이용할 수 있다. 각 결과 값과 상기 결과 값의 평균을 산출하여 이용할 수 있다. 또는 각 항목별로 가중치를 부여하여 가중치가 부여된 결과 값의 평균을 산출할 수도 있다.
- [0086] 520 단계는 상기 산출된 결과 값의 평균으로부터 기술가치 순위를 선정하는 단계이다.
- [0087] 상기 산출된 결과 값의 평균을 이용하여 각 특허 문서의 기술가치 순위를 선정한다. 상기 산출된 결과 값의 평균이 낮을 수록 높은 기술가치 순위를 부여할 수 있다.
- [0088] 기술가치 순위는 다음 표와 같이 나타낼 수도 있다.

표 3

특허문서	기간	청구항수	IPC코드	키워드	군집중심점	중개중심점	평균	기술가치 순위
1504375	13	8	4	3	1	1	5	1
1596313	3	20	10	18	1	1	8.83	2
1531456	13	5	4	18	10	10	10	3
...	...	...	...	...	...	...	...	...

- [0090] 도 6은 본 발명의 실시예에 따른 특허분석방법의 실시예로 610 단계를 더 포함할 수 있다.
- [0091] 610 단계는 상기 사회네트워크분석 결과, 상기 추출된 키워드, 또는 상기 선정된 중요 특허들 중 하나 이상을 사용자에게 제공하는 단계이다.
- [0092] 보다 구체적으로, 110 단계 내지 140 단계의 결과를 사용자에게 제공하는 단계이다. 사회네트워크분석을 통해 도출된 중요 특허, 중요 키워드 추출을 통해 추출된 중요 키워드, 기술가치 순위를 통해 추출된 중요 특허, 또는 k-means 분석을 통해 도출된 중요 기술군집의 대표기술 등을 사용자에게 제공하는 단계이다. 상기 사용자에게 제공되는 정보들을 테이블로 생성하여 사용자에게 화면으로 출력할 수 있다.
- [0093] 도 7 내지 도 10은 본 발명의 실시예에 따른 특허분석의 과정을 나타낸 것이다.
- [0094] 기술가치와 시장가치를 결합한 통합된 기술평가모형으로써, 평가대상이 되는 기술들의 특허를 이용한다. 특허는 IPC코드, 청구항, 출원일, 요약, 발명자, 명칭 등과 같은 정보들을 포함하고 있다. 대부분의 기술이전의 경우, 특허들은 대학이나 연구소에서 회사나 기업으로 이전된다. 기술평가와 추가적인 기술평가를 위한 특허군집을 위해 특허를 이용한다. 기술 그룹들은 유사한 기술들을 포함하는 특허들로 구성 된다. 그런 이유로 기술평가모형은 특허와 특허군집에 대한 평가를 기반으로 한 2가지 기술평가방법으로 이루어진다. 특허를 평가하기 위해 우

리는 기계학습방법과 통계학과 같은 다양한 정량적인 방법을 이용해 특허문서를 분석한다.

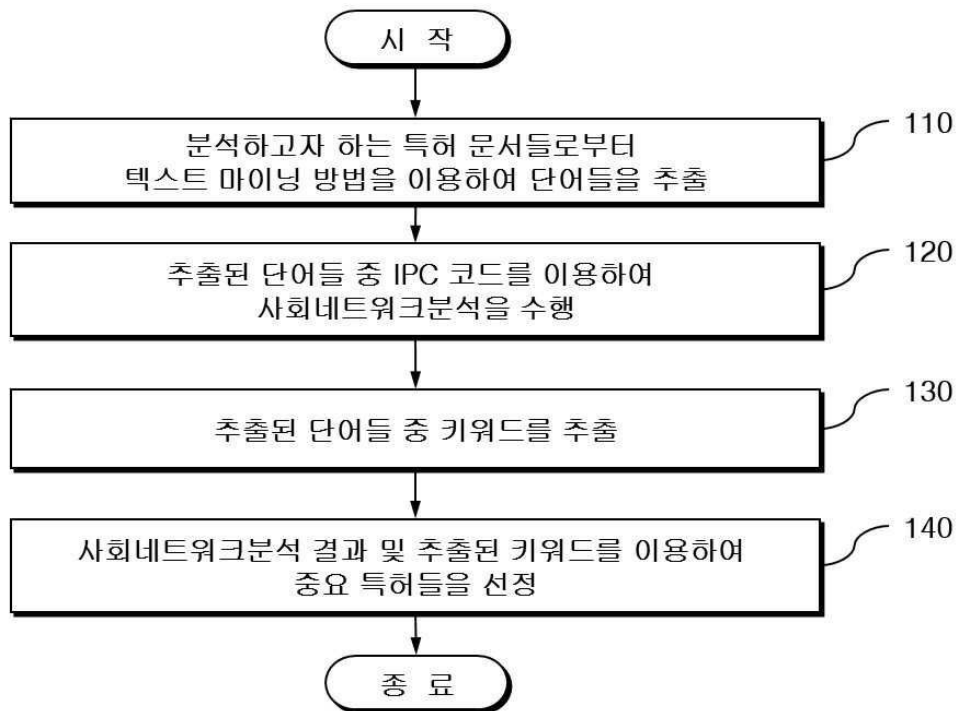
- [0095] 도 7은 정량적 특허 분석을 나타낸 것으로, 특허는 논문이나 기사보다 개발된 기술에 대한 더 다양하고 복잡한 정보를 포함하고 있다. 그러므로 특허문서는 기술분석과 예측을 위한 고급정보이다. 기술경영분야에서는 대부분 정량적인 방법의 특허 분석 결과를 이용한다. 정량적인 특허 분석이란 도 7과 같이 기계학습 방법과 통계학을 기반으로 한 수학적인 Tool을 이용해 특허 문서를 분석하는 것이다.
- [0096] 분산, 평균, 빈도 표와 같은 기술 통계학을 이용하여 IPC 코드, 청구항 수 등의 대표 변수를 이용하여 검색된 특허문서들을 요약할 수 있다. 하나의 IPC 코드는 4개의 등급 (1: Section, 2: Class, 3: Subclass, 4: group)으로 구성되어 있다. IPC코드의 4번째 등급까지 이용할 경우 너무 세세하게 분류되기 때문에 Subclass까지 이용할 수 있다. 회귀 모형과 연관 규칙을 이용하여 특허간의 의미 있는 관계를 도출하였다. 성능이 좋은 연관 규칙 모형인 사회 네트워크 분석을 이용하였다. 군집과 분류는 정량적인 분석을 위한 또 다른 접근법이다. 예를 들어 우리는 K-means 알고리즘을 이용한 특허 군집화를 통해 기술 분석을 수행할 수 있다. 특허 문서 분석을 위해 정량적인 분석 방법을 이용하였다.
- [0097] 기술평가는 기술 라이선스, 기술 마케팅, 영업 비밀, 기술 이전 등을 위한 경제적 가치를 기반으로 한 기술의 측도로 사용되었다. 기술 통계학, 텍스트 마이닝, 기술 군집화, 사회 네트워크 분석을 결합한 정량적인 특허 분석을 새로운 기술평가방법으로 이용한다.
- [0098] 도 8은 텍스트 마이닝을 나타낸 것으로, 특허 데이터베이스로부터 수집된 특허 문서들은 그림, 인용도, 청구항 수, 요약, 발명의 명칭, 출원일, 발명자와 같은 그림, 숫자, 텍스트를 포함한 여러 다른 종류들로 이루어져 있다. 이 데이터는 기계학습 알고리즘 또는 통계적 분석 방법을 분석할 수 없다. 이 문제를 해결하기 위해 텍스트 마이닝을 이용한 전처리 과정을 통해 특허 문서를 기계학습 방법과 통계 분석에 적합한 형태로 변환한다. 텍스트 마이닝은 텍스트 문서로부터 의미있는 단어를 추출하는 데이터 마이닝 기법이다. 정량적인 특허분석을 위한 구조화된 데이터를 생성하기 위해 검색된 특허 문서들로부터 키워드들을 추출한다. 도 8은 수집된 특허 문서들의 변환을 위한 텍스트 마이닝 기법을 이용한 전처리 과정을 보여준다.
- [0099] 먼저 분석 대상이 되는 기술과 관련된 키워드 검색식을 이용해 특허 데이터베이스로부터 해당 특허들을 검색한다. 검색된 특허들 중에 가치 있고 의미 있는 특허들을 선별하여 텍스트 마이닝을 이용한 전처리 과정을 수행한다. 필터링 된 특허 문서들로부터 텍스트 마이닝 기법을 이용해 문서 안에 모든 단어들을 추출한다.
- [0100] Step 1. Preprocessing
- [0101] Preparing filtered patent documents
- [0102] Importing patent data
- [0103] Creating structured data
- [0104] Constructing corpus, text document collection
- [0105] Extracting keywords from corpus
- [0106] Building term(keyword)-document matrix for analysis
- [0107] Step 2. Analyzing
- [0108] Finding association rules between patent or keywords
- [0109] Modeling dependencies between technologies
- [0110] Clustering patents using similarity and silhouette measures
- [0111] 텍스트 마이닝 과정은 위와 같이 전처리 과정과 분석 과정 2단계로 이루어 진다. 전처리 과정을 통해 접속사와 같은 의미 없는 단어들을 제거하고 통계 분석을 위한 구조화된 데이터로써 단어-문서 행렬을 생성한다. 단어-문서 행렬의 행과 열은 각각 특허 문서와 단어들로 이루어지고 각 셀은 특허 문서 안에 단어들의 출현 빈도를 나타낸다. 이 단어-문서 행렬을 이용해 기계학습 방법과 통계 분석을 수행할 수 있다.
- [0112] 도 10은 사회 네트워크 분석을 나타낸 것으로, 사회 네트워크 분석은 네트워크 안에 노드들 간의 사회적 연결관계를 분석한다. 사회 네트워크 분석은 정량적인 특허 분석에 사용될 수 있다. 사회 네트워크 분석의 입력 데이터로 각 특허들의 기술을 대표하는 키워드를 사용할 수 있다. 도 10은 정량적인 특허 분석을 위한 본 발명의 사

회 네트워크 분석 방법을 보여준다.

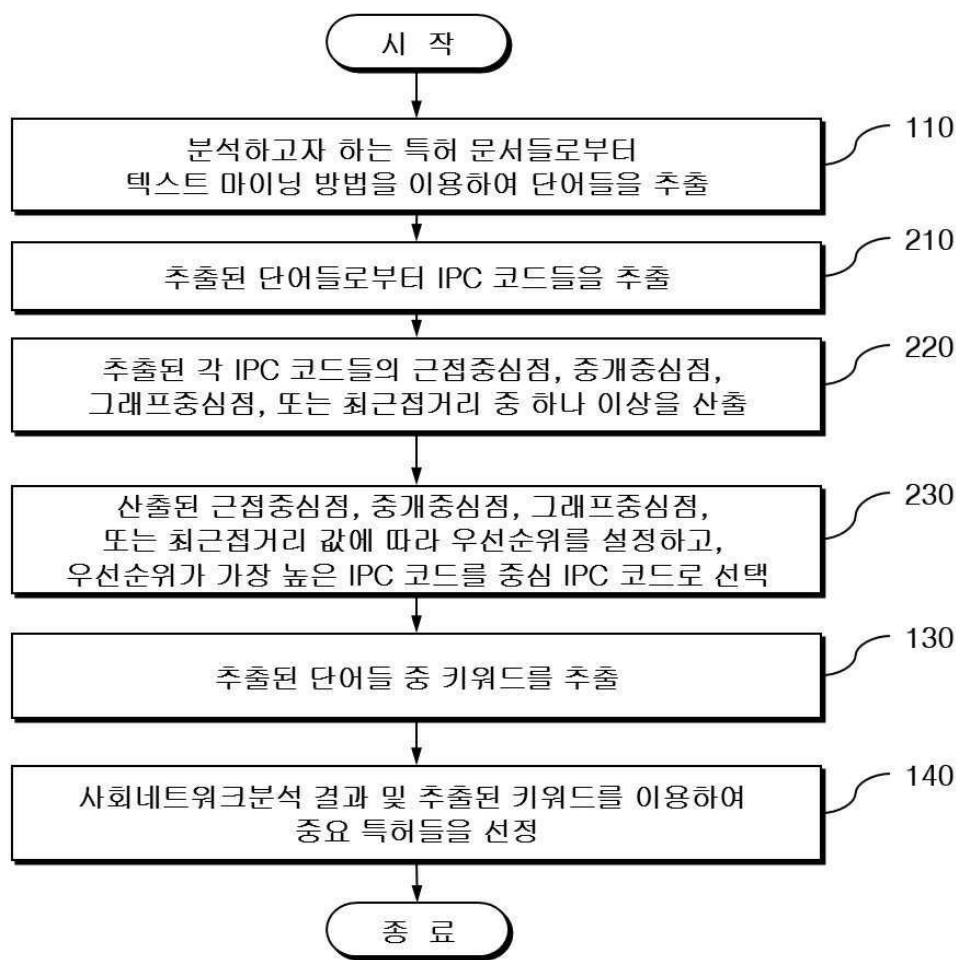
- [0113] 텍스트 마이닝 결과를 이용해 만든 특허 문서-키워드 행렬을 사회 네트워크 분석에 입력 데이터로 이용하기 위해 키워드-특허 문서 행렬로 변환하였다. 그리고 특허들 간의 상관 계수를 계산한다. 이 결과는 사회 네트워크 분석에서 인접 행렬로 사용되고 가치 있는 특허들을 선택하기 위한 사회 네트워크 분석 그래프를 그리는데 이용된다. 사회 네트워크 분석 결과의 성능 평가를 위해 근접 중심성과 중개 중심성 두 가지 측도를 고려한다.
- [0114] 사회 네트워크 분석은 vertex (or node)와 edge (or connection) 또는 edge들의 분석에 의한 vertex들 사이의 관계설명 등 총 2가지로 구성되어있다. 근접중심점(Closeness centrality, CC), 중개중심점(Betweenness centrality, BC), Graph centrality, shortest distance를 목표기술(target technology)의 대표기술 선정에 위한 기준으로 사용한다.
- [0115] 도 10은 기술 군집화를 나타낸 것으로, 군집화는 거리측도를 이용해 유사한 그룹으로 모든 객체들을 나누는 것이다. 특허 문서는 청구항, 요약, 발명의 명칭과 같은 많은 요인(변수)를 갖고 있다. 이 변수들을 기반으로 모든 수집된 문서들을 군집화시킬 수 있다. 특허 군집화의 결과를 이용해 기술의 특징과 적용을 기준으로 한 기술 비교가 가능하다. 수집된 특허 문서를 이용해 특허 군집화를 수행한다. K-means 군집화를 수행하기 위해서는 군집 개수를 분석자가 정해줘야 한다. 실루엣 측도를 이용해 최적의 군집 수를 구하여 군집화를 수행한다. 분석 결과인 특허 군집을 기술평가에 이용한다.
- [0116] 대부분의 전통적인 기술평가 프로세스는 전문가의 지식에 의존해 기술 또는 특허에 대한 평가를 해왔다. 전통적인 기술평가는 전문가에 따라 주관적인 결과를 얻을 수 밖에 없었다. 그러나 기업들은 기술 이전, 기술 파이낸싱 등에서 리스크를 줄이기 위해 객관적인 결과가 필요하다. 이를 해결하기 위하여 본 발명의 실시예에 따른 특허분석방법은 텍스트 마이닝 기법, 기계학습 방법, 통계적인 분석 방법을 이용한 객관적인 기술평가 모형을 구축하였다. 또한 기술평가 모형에 특허를 입력데이터로 이용해 기술 이전 및 기술 파이낸싱 등에서 리스크를 줄이기 위한 객관적인 기술평가 결과를 제공할 수 있다.
- [0117] 본 발명의 실시예는 특허문서에 대한 사회 네트워크 분석 및 키워드 추출을 수행하여 특허문서를 분석하는 하나 이상의 처리부(프로세서) 및 처리부에서 산출되는 결과들을 저장하는 하나 이상의 저장부(데이터베이스)를 포함할 수 있다. 나아가, 정보를 사용자에게 제공하는 영상출력부(디스플레이)를 더 포함할 수 있다.
- [0118] 본 발명의 실시예들은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 매체에 기록되는 프로그램 명령은 본 발명을 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(magnetic media), CD-ROM, DVD와 같은 광기록 매체(optical media), 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media), 및 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다. 상기된 하드웨어 장치는 본 발명의 동작을 수행하기 위해 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.
- [0119] 이상과 같이 본 발명에서는 구체적인 구성 요소 등과 같은 특정 사항들과 한정된 실시예 및 도면에 의해 설명되었으나 이는 본 발명의 보다 전반적인 이해를 돕기 위해서 제공된 것일 뿐, 본 발명은 상기의 실시예에 한정되는 것은 아니며, 본 발명이 속하는 분야에서 통상적인 지식을 가진 자라면 이러한 기재로부터 다양한 수정 및 변형이 가능하다.
- [0120] 따라서, 본 발명의 사상은 설명된 실시예에 국한되어 정해져서는 아니되며, 후술하는 특허청구범위뿐 아니라 이 특허청구범위와 균등하거나 등가적 변형이 있는 모든 것들은 본 발명 사상의 범주에 속한다고 할 것이다.

도면

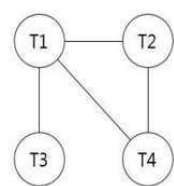
도면1



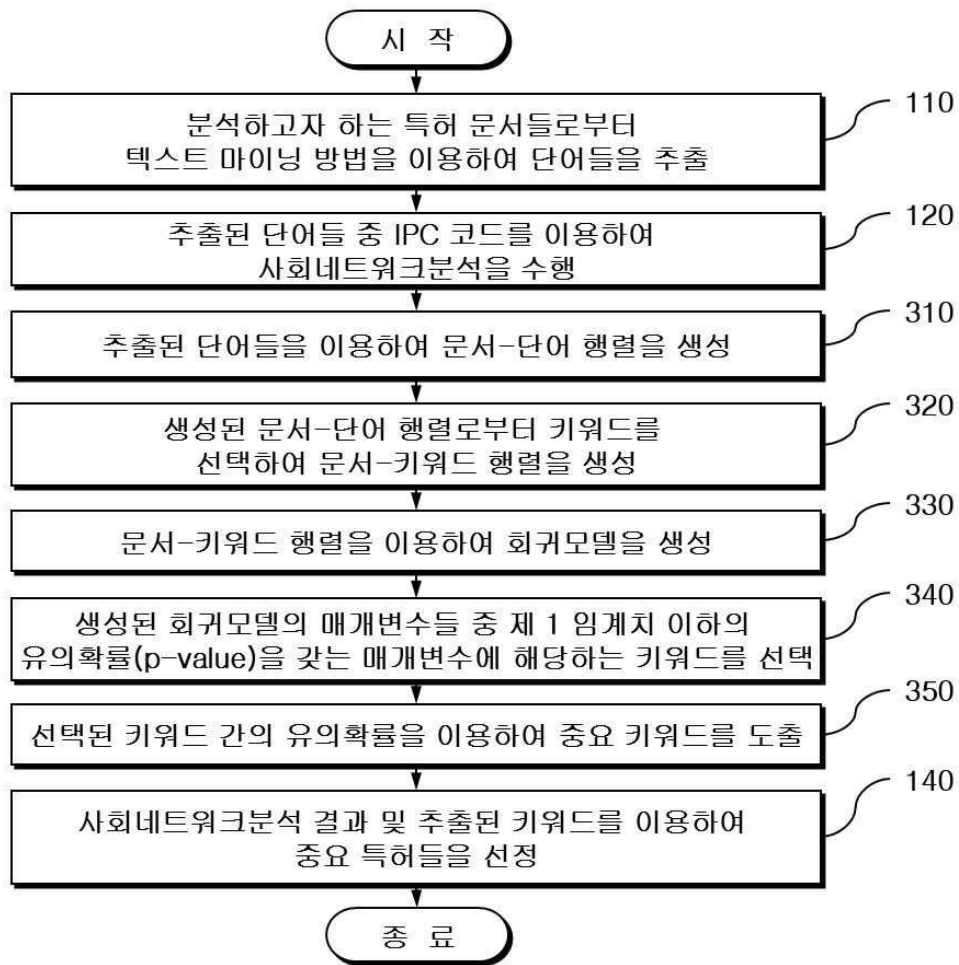
도면2



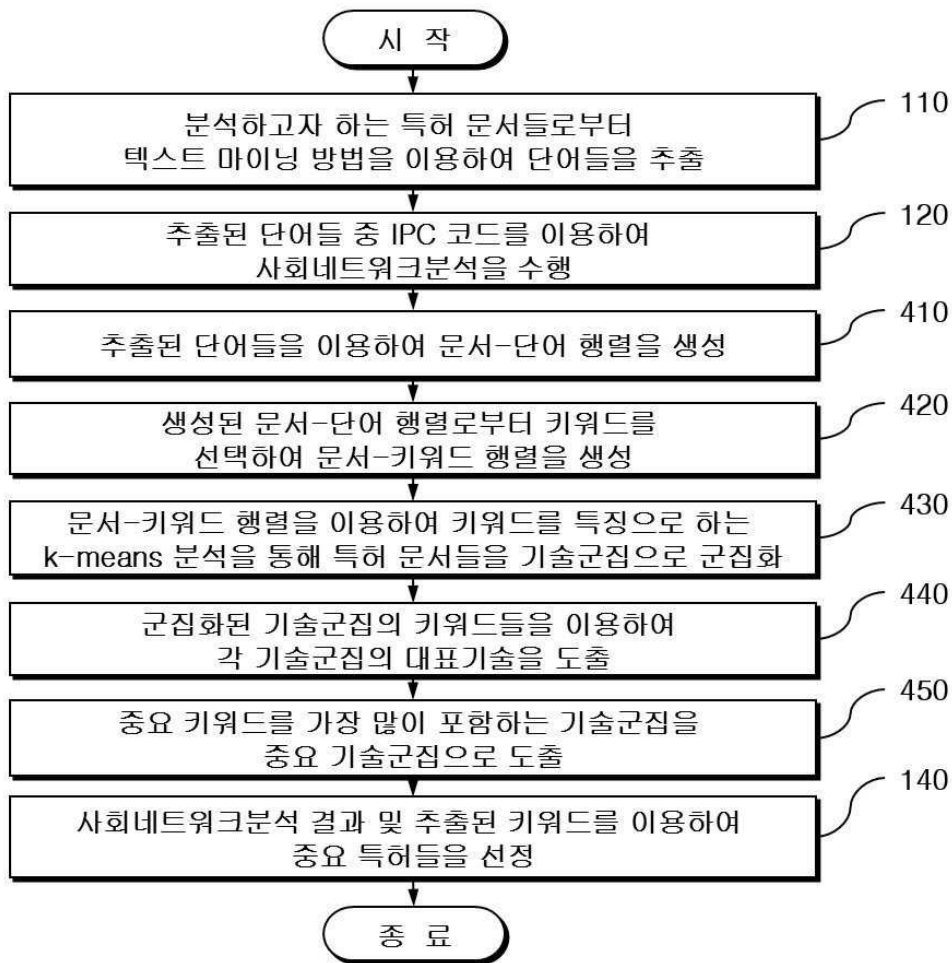
도면2a



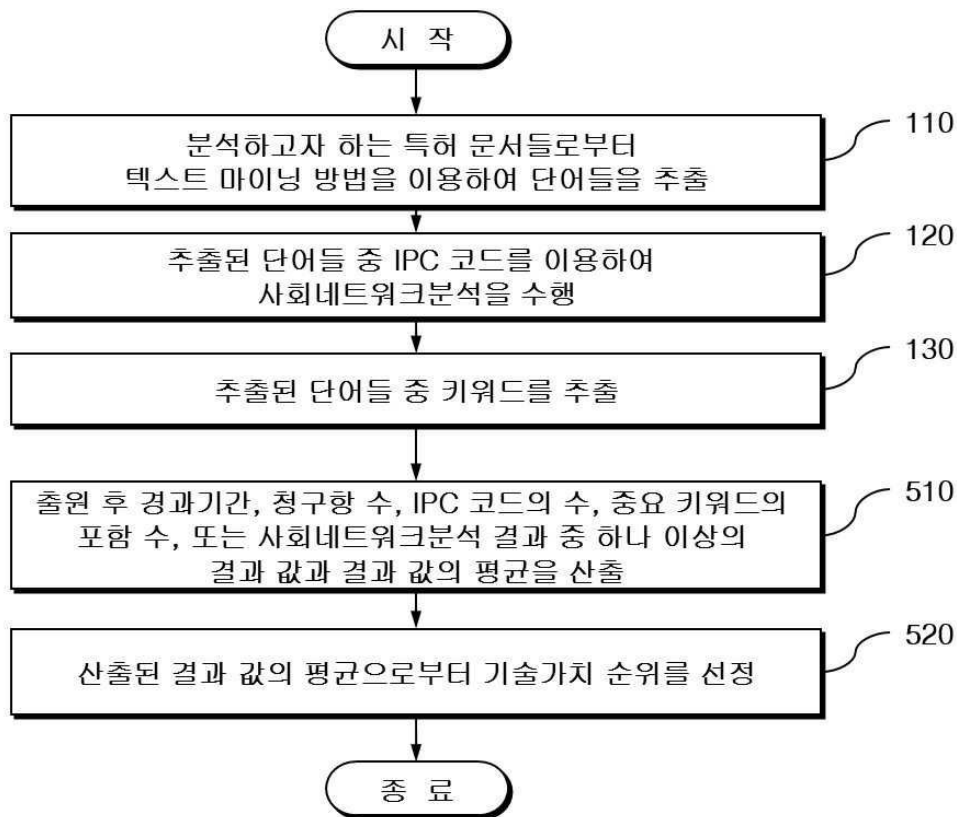
도면3



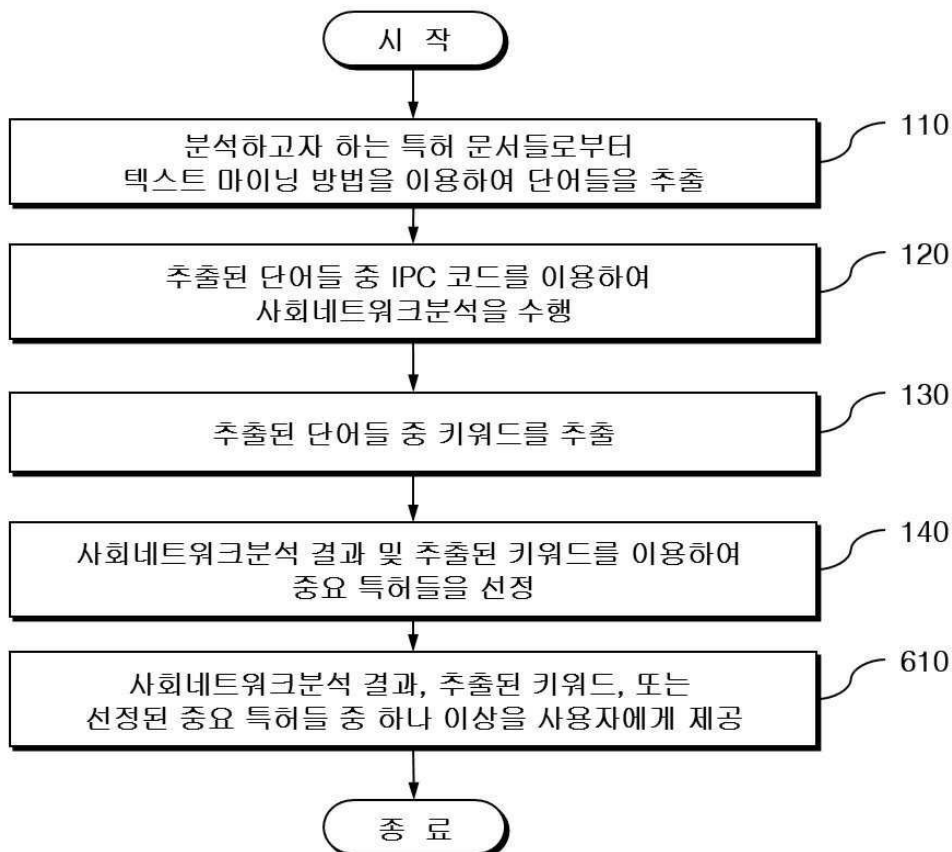
도면4



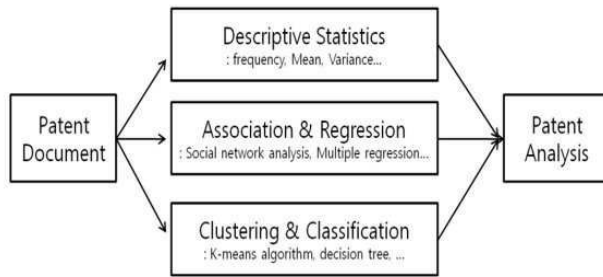
도면5



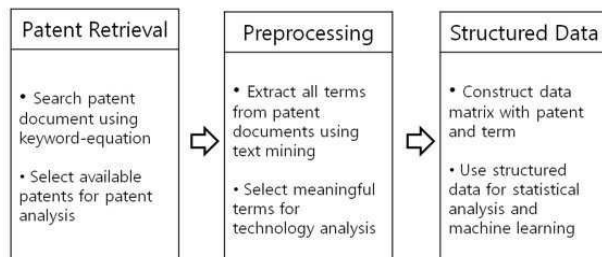
도면6



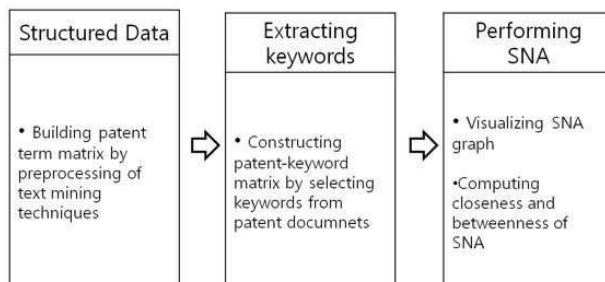
도면7



도면8



도면9



도면10

